17th International Conference on the Applications of Computer
Science and Mathematics in Architecture and Civil Engineering
K. Gürlebeck and C. Könke (eds.)
Weimar, Germany, 12–14 July 2006

THE STRUCTURE OF ROAD TRAFFIC SCENES AS REVEALED BY UNSUPERVISED ANALYSIS OF THE TIME AVERAGED OPTICAL FLOW

U. Knauer*, T. Dammeier, and B. Meffert

*Humboldt-Universität zu Berlin, Institut für Informatik
Unter den Linden 6, 10099 Berlin, Germany
E-mail: { knauer | dammeier | meffert }@informatik.hu-berlin.de

Keywords: Optical flow, Scene model, Segmentation.

Abstract. *The Lucas-Kanade tracker has proven to be an efficient and accurate method for calculation of the optical flow. However, this algorithm can reliably track only suitable image features like corners and edges. Therefore, the optical flow can only be calculated for a few points in each image, resulting in sparse optical flow fields. Accumulation of these vectors over time is a suitable method to retrieve a dense motion vector field. However, the accumulation process limits application of the proposed method to fixed camera setups. Here, a histogram based approach is favored to allow more than a single typical flow vector per pixel. The resulting vector field can be used to detect roads and prescribed driving directions which constrain object movements. The motion structure can be modeled as a graph. The nodes represent entry and exit points for road users as well as crossings, while the edges represent typical paths.*

1 INTRODUCTION

The application of optical systems in traffic monitoring has the potential for an automatic traffic data generation. Such systems are not limited to aquire information which is related to cross sections or to lanes as it is typical for other techniques like induction loops or floating car data. Therefore, optical systems offer a chance of further optimizing traffic flow on intersections, to identify stalled vehicles or accidents, or to give a forecast of information obtained from one intersection to the next. Research and development in computer science contribute to this field by the design of algorithms and hardware for signal processing and image analysis [1].

Object movement in urban scenarios is not a random process. In fact, it is strongly restricted by the structure of the environment. Objects move within defined areas (streets, sidewalks), avoid obstacles and are affected by traffic regulations.

Although the movement of pedestrians is less restricted compared to vehicles, they tend to restrict themselves by the way they choose their paths. Even on public spaces where people are free to walk anywhere, there are invisible paths used by the majority of people. These paths are the optimal connections between points of interest as well as entry and exit zones. They are subconsciously chosen in the same way by most individuals (see Fig.1).

As a result the direction of motion at every position is limited to only a few alternatives. The same applies to the speed of motion but here the variety of values is less restricted. Usually there is no lower speed limit for object motion, only maximum values due to physical limitations of pedestrians and speed limits for vehicles.

Observing such a scene with a fixed camera one would expect the existence of one or more motion vectors for every image pixel, describing the typical object motions for this particular point.

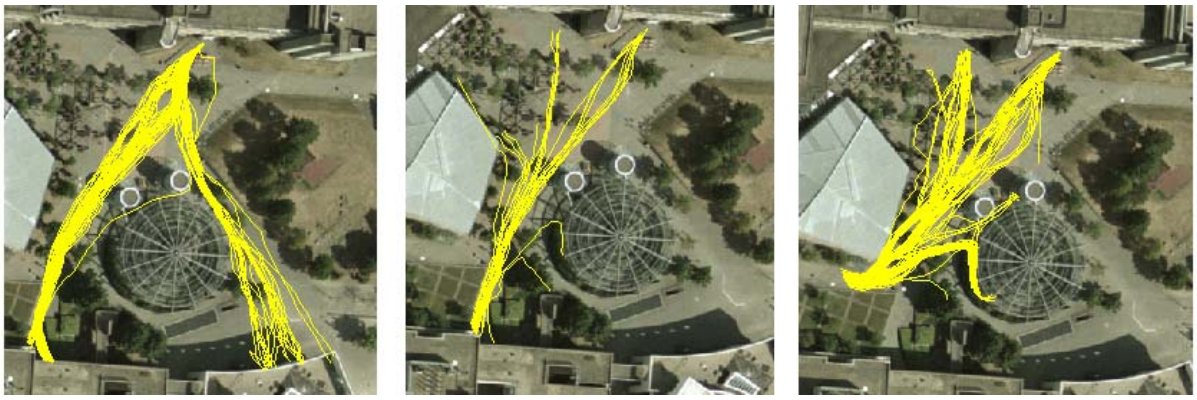


Fig. 1 Common paths as revealed by pedestrian trajectories

If it is possible to retrieve this information for every point of the scene then the result will be a valuable description of the motion situation in the scene and contains the structure of the scene as revealed by object movement. Here a method for automated retrieval of this information by time averaging the local optical flow information for each pixel is proposed. Different measures are described which are derived from the optical flow. An example is given how the data can be analysed in order to create a model of the scene. It is important to note that this is done without tracking any of the objects in the scene itself. So, this approach does not depend on object segmentation and object tracking.

The scope of the method is not solely the acquisition of traffic data but the development of a supplementary component which supports the set up of such systems by identifying the regions of interest (e.g. lanes) as well as their properties (driving directions, speeds, etc.).

2 CALCULATION OF THE OPTICAL FLOW

To calculate the optical flow between successive video frames the well known combination of feature selection as introduced by Shi and Tomasi [2] and the algorithm of Lucas and Kanade for feature tracking [3] is used. Feature selection finds image blocks which are believed to allow the accurate estimation of the optical flow translation vector. The Shi-Tomasi algorithm utilizes the smallest eigenvalue of an image block as criterion to ensure the selection of features which can be tracked reliably by the Lucas-Kanade Tracking algorithm.

This algorithm matches the selected image blocks with blocks in the next frame using an efficient gradient descent technique. A pyramidal implementation of this algorithm is used to deal with larger feature displacements by avoiding local minima in a coarse to fine approach [4]. This combination has proven to allow fast and reliable computation of optical flow information [5,6].

To summarize, the feature selection process is controlled by:

- an optional upper limit for the number of significant image blocks,
- feature quality, defined by the ratio between the smallest eigenvalue of a block compared to the most significant block (see [2] for details),
- the minimum required distance between two selected significant blocks, and
- optionally a motion mask.

The first parameter was introduced for computational reasons. It ensures that only a maximum number of significant blocks have to be tracked by the Lucas-Kanade algorithm.

The second parameter, feature quality, is linked to the gradient information of each block which is supposed to correspond to the uniqueness of the image block for matching. The parameter automatically adapts to illumination changes because the maximum value is determined by the current frame.

Feature quality and the third parameter, minimum distance, are different ways to control the number and quality of selected blocks. Increasing the lower bound for feature quality will result in selecting fewer blocks for tracking and a higher feature distance forces distributed and non overlapping blocks.

A thresholded difference image is used as a mask to restrict feature selection and tracking to areas where relevant movement exists. This serves as a replacement for an object mask and is much easier to retrieve and sufficient for the task. A fixed threshold is adequate since the aim is to divide image noise from intensity value changes caused by moving objects. In general the difference values caused by motion are large enough to find a suitable threshold which suppresses noise without loosing too much of the object.

The parameters which control the performance of the pyramidal Lucas-Kanade algorithm are:

- the block size,
- the number of resolution levels, and
- some termination criteria.

The block size and the number of resolution levels are determined by the image size and the expected block displacement between two frames of an image sequence. The termination criteria are the maximum number of iterations for the gradient descent approach and an accuracy requirement for block matching to allow early termination.

Finally, the results are filtered for relevance and quality:

- To reduce the number of mismatched motion vectors the RMSE of each pair of blocks is evaluated and vectors with high error values are discarded.
- Speed thresholds are used to focus evaluation to certain speed ranges and to suppress very small movements which are often caused by image noise and unrealistic large displacements.

Fig. 2 shows motion vectors for initially selected significant blocks and the remaining vectors after relevance filtering. Vectors discarded by RMSE are drawn in red.

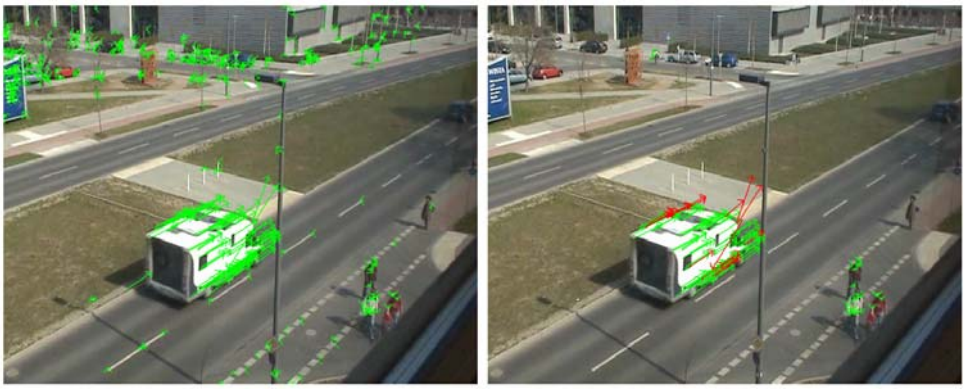
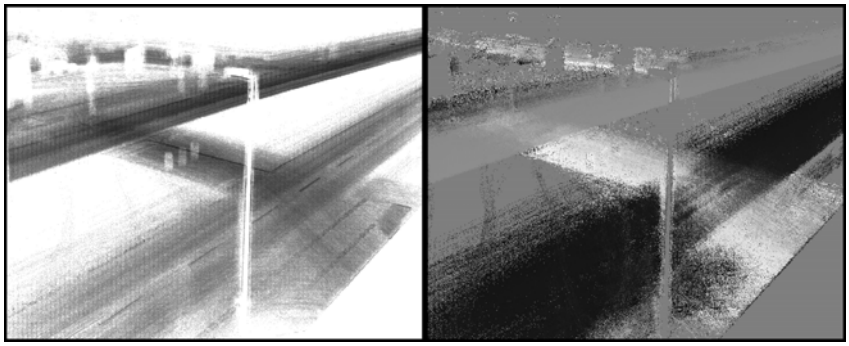


Fig. 2 Optical flow for significant image points before and after relevance filtering

3 ACCUMULATION PROCESS

The extracted motion vectors at each pixel are collected and averaged over time. This way it is possible to retrieve a time averaged optical flow field consisting of an average motion vector for each image pixel. The number of retrieved vectors at each point is used as a measure for the motion activity and for the statistical quality of the mean values (Fig. 3a). However, the mean value does not necessarily represent the movement situation e.g. at cross sections. Therefore, a histogram based approach is favored to allow more than one typical direction for each image pixel. By storing a histogram with 12 bins, each covering an angle of 30 degree, information about the distribution of movement directions is obtained.

The different gray values in Fig. 3b illustrate the average angles of the motion vectors. As can also be seen in Fig. 3 the generated fields suffer from problems like perspective distortions and partial occlusions since the measurable two dimensional block displacement is just the result of a projection from motion in a three dimensional world. These problems can be minimized by camera calibration and an appropriate camera position.



(a) (b)

Fig. 3 Frequency of motion (a) and average direction (b) at each pixel

As described before, only a limited number of motion vectors can be obtained for each pair of consecutive images. Therefore, a sufficient observation time is required to produce a dense and reliable averaged optical flow field. In general, the field quality will increase as long as the traffic situation in the observed scene remains constant. First experiments indicate that the processing of at least 50,000 frames is necessary. Fig. 4 illustrates the number of motion vectors which have been obtained at each location.

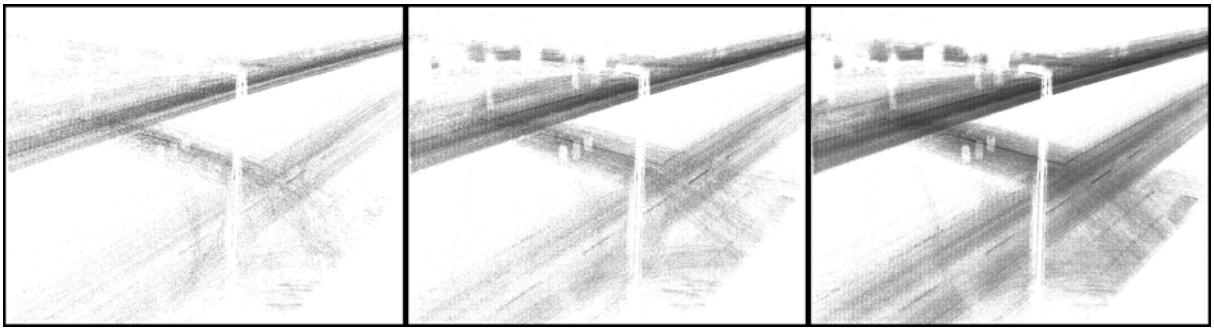


Fig. 4 Frequency of movement per pixel after 10000, 25000 and 50000 frames, respectively

4 FIELD SEGMENTATION

This section gives an example, how information about the structure of the scene can be obtained from the motion vector field with standard image processing techniques. As described before, the frequency of movement is recorded for twelve different directions. In Fig. 5a this frequency is shown for a single direction as an intensity image.

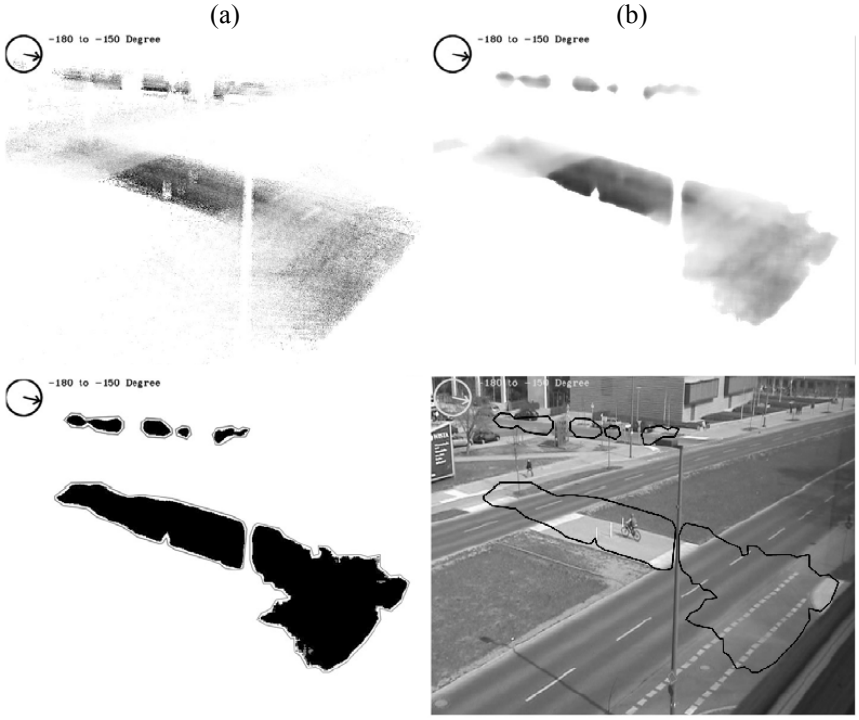


Fig. 5 Frequency of movement for a single direction: (a) raw, (b) median-filtered, (c) binarized, and (d) mask superimposed on the input image

The intensity images (Fig. 5a) are median-filtered (Fig. 5b) and binarized (Fig. 5c). This procedure results in twelve binary masks.

A classical approach to analyze the image structure is the computation of its skeleton. However, different algorithms and/or starting points lead to different results. Therefore, the main axis of each region instead of its skeleton is calculated. Because each binary image is derived from the direction of movement it is easy to select the correct axis from the result of a Karhunen-Loeve transform (KLT).

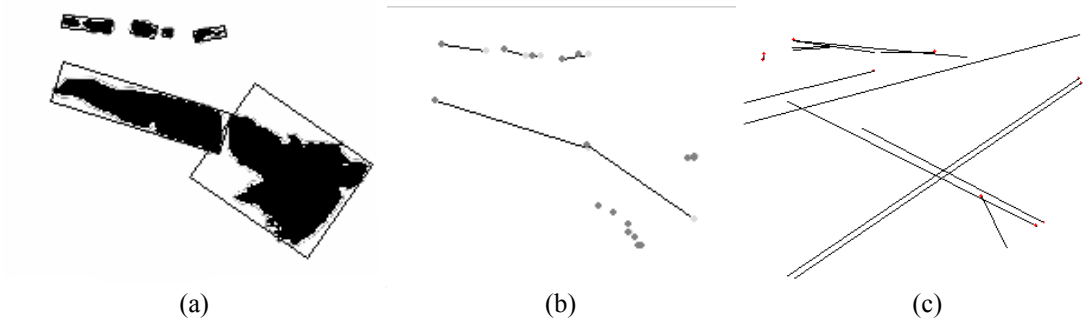


Fig. 6 Finding principal components (a), extracted main axes (b), and fusion of the results for twelve masks (c)

Fig. 6 shows the result of the two-dimensional KLT for a single image and the fusion of all images. To improve the result, adjacent lines are merged. Note, that Fig. 6c provides information about the direction of movement.

5 TRAFFIC DATA GENERATION

So far, the scope of the proposed method has been the long-term acquisition of motion vectors in order to derive typical paths without including too much prior knowledge. However, if the exterior orientation of the camera is known then the paths as well as the speed information can be mapped into a real world coordinate system. Fig. 7 shows an example.

Fig. 7 Camera field of view and aerial image of the same scene

The projective transformation which maps the camera field of view to the georeferenced aerial photo can be calculated from at least four pairs of corresponding points inside both images. Fig. 8 shows a screenshot from the control point selection tool which is part of the Matlab Image Processing Toolbox and can be used for this task.

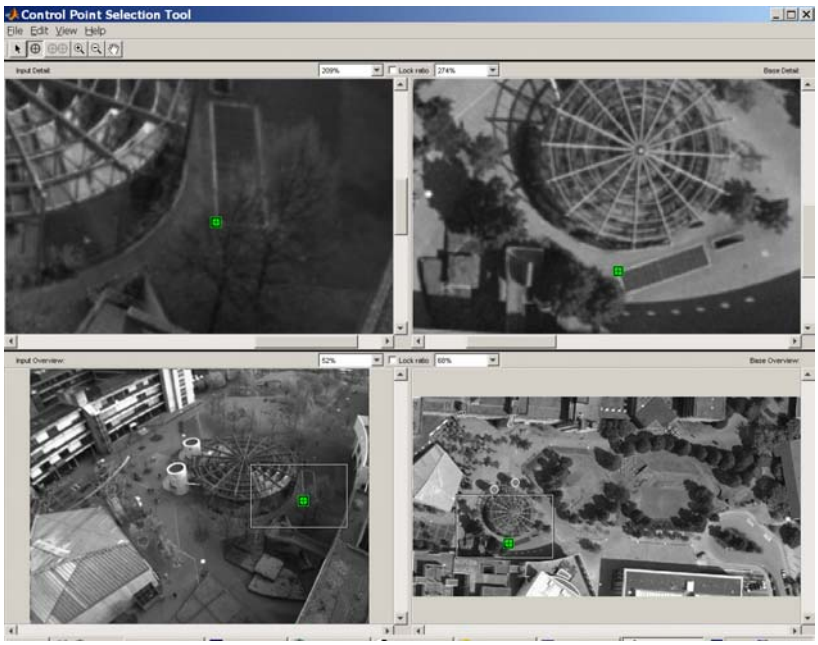


Fig. 8 Matlab control point selection tool

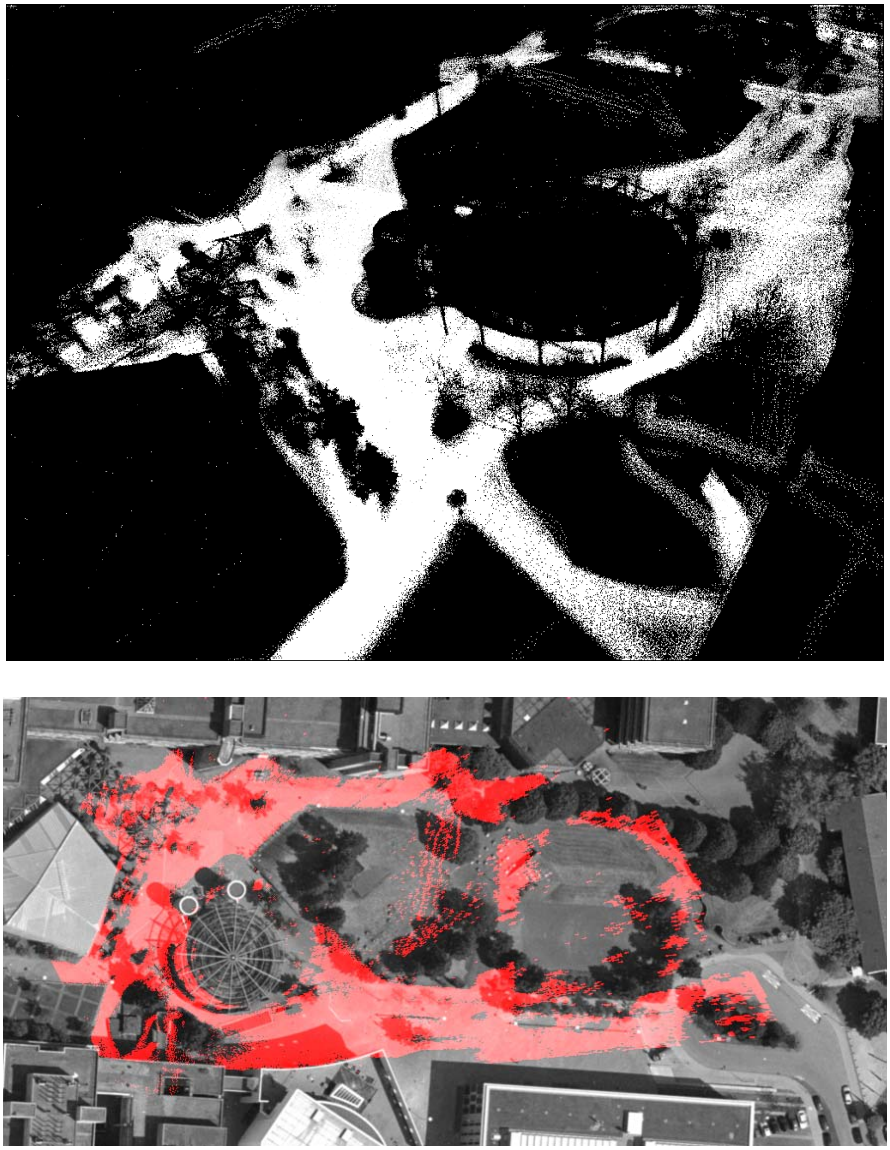


Fig. 9 Result of the projective transform

The speed information combined with the most frequent direction at each image point fixes the coordinates of a second point. Both points can be mapped into the real world coordinate system applying the projective transform. The distance of the mapped points is used to calculate an estimate for the typical ground speed.

6 CONCLUSION

Vector fields which reflect the motion situation of an observed scene can be retrieved by accumulation of local optical flow information. Optical flow vectors thereby used are not expected to be free of error, but due to the nature of averaging the effect of false vectors in the averaged motion field should be minimized.

Since block matching is done only between consecutive frames it is nearly not affected by changes in illumination conditions. Even large and fast illumination changes result in only relative small interframe intensity value differences, not influencing the matching algorithm.

However, if bad illumination conditions result in images with low overall contrast the number of significant blocks will decrease, which increases the necessary observation time in order to retrieve meaningful averaged fields.

For the construction of a scene model a similar approach is presented in [7]. There a scene model is derived from object trajectories. The major drawback of such an approach is the dependence on object segmentation and tracking, that is still limited in outdoor scenes. Therefore, the main advantage of the proposed method is the stable tracking of low level features and its stability due to averaging. In general, uncertainty of information about object positions and sizes restricts the quality of any tracker. Kalman or particle filters are typically used to overcome some of the limitations. These kinds of filters are based on statistical models of motion. Such models are typically initialized with a default and adapt to the movement of the object. Therefore, they rely on temporal rather than spatial information. The motion vector field provides information about speed and direction at any pixel where motion typically occurs. It can easily be used as a supplementary component for motion prediction to improve tracking results. Such a component may use off-line results which are acquired during a set up phase or continuously updated data.

Future work will concentrate on the application of the scene model for camera self localization and estimation of its exterior orientation, as well as on the refinement of the algorithm.

ACKNOWLEDGEMENT

This work has been financially supported by the IBB program ProFit.

REFERENCES

[1] Meffert, B.; Blaschek, R.; Knauer, U.; Reulke, R.; Schischmanow, A.; Winkler, F.: Monitoring Traffic by Optical Sensors, Proc. 2nd International Conference on Intelligent Computing and Information Systems, Cairo, Egypt, (2005), pp. 9-14.

[2] Shi, J.; Tomasi, C.: Good Features to Track. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, (1994), pp. 593-600.

[3] Lucas, B. D.; Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. of 7th International Joint Conference on Artificial Intelligence, (1981), pp. 674-679.

[4] Bouguet, J.-Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker. Part of OpenCV Documentation.

[5] Barron, J. L.; Fleet, D. J.; Beauchemin, S. S.; Burkitt, T. A.: Performance of optical flow techniques. In: International Journal of Computer Vision 12, Vol. 1 (1992-02), pp. 43-77.

[6] Baker, S. ; Matthews, I.: Lucas-Kanade 20 years on: A unifying framework: Part 1. Technical Report CMU-RI-TR-02-16, CMU Robotics Institute, (2002).

[7] Makris, D.; Ellis, T.J. : Automatic Learning of an Activity-Based Semantic Scene Model. In: Proc. IEEE International Conference on Advanced Video and Signal Based Surveillance, (2003), pp. 183-188.